

METHODS AND SYSTEMS FOR ESTIMATING A CHANNEL RESPONSE BY APPLYING BIAS TO AN INITIAL CHANNEL ESTIMATE

BACKGROUND OF THE INVENTION

This invention relates to digital communications systems and methods, and more particularly to systems and methods for estimating a channel response from training symbols that are received over a channel.

5 Digital communications systems and methods are widely used for voice, image and/or data communications. As is well known to those having skill in the art, when communicating wirelessly, the time-varying multipath fading of wireless channels may make reliable transmission difficult. In efforts to reduce the effect of such time-varying channels, it is known to transmit a sequence of training symbols, which are known to both the transmitter and the receiver. The known training symbols may be provided within each burst or selected bursts of transmitted symbols, along with the payload symbols that are unknown to the receiver. Based on these known training symbols, a channel response may be estimated. The channel estimate can then be used to demodulate the unknown payload symbols.

15

SUMMARY OF THE INVENTION

A channel response may be estimated from training symbols that are received over a channel, according to some embodiments of the present invention, by determining an initial channel estimate from the training symbols that are received over the channel, and adding bias to the initial channel estimate to obtain a biased channel estimate. The biased channel estimate may then be used to demodulate a signal that is received over the channel. In some embodiments, bias is added by transforming the initial channel estimate according to a transform matrix, setting at least one tap of the transformed initial channel estimate to zero, and inverse transforming the transformed initial channel estimate with at least one tap set to zero, to obtain the biased channel estimate. The transforming, setting at least one tap and inverse transforming may be repeatedly performed.

20
25

Accordingly, some embodiments of the present invention can add bias to an initial channel estimate to obtain a biased channel estimate that reduces error in the estimated channel response compared to the initial channel estimate. It will be understood that, although the above description has focused primarily on methods of
5 estimating a channel response, analogous systems and wireless receivers that employ similar techniques to estimate any other parameters also may be provided.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 illustrates a conventional format of a burst of transmitted symbols in
10 a conventional GSM/EDGE system.

Figure 2 is a block diagram of a conventional wireless receiver.

Figure 3 is a block diagram of systems, methods and receivers for estimating a channel response from training symbols that are received over a channel according to some embodiments of the present invention.

15 Figure 4 is a flowchart of operations that may be performed to estimate a channel response according to other embodiments of the present invention.

Figure 5 is a block diagram of other embodiments of the present invention.

Figure 6 is a flowchart of operations that may be performed according to other
embodiments of the present invention.

20 Figures 7 and 8 graphically illustrate gains that may be achieved by embodiments of the present invention.

DETAILED DESCRIPTION

The present invention now will be described more fully hereinafter with
25 reference to the accompanying drawings, in which embodiments of the invention are shown. This invention may, however, be embodied in many different forms and should not be construed as limited to the embodiments set forth herein. Rather, these embodiments are provided so that this disclosure will be thorough and complete, and will fully convey the scope of the invention to those skilled in the art. Like numbers
30 refer to like elements throughout. It will be understood that when an element is referred to as being "connected" or "coupled" to another element, it can be directly connected or coupled to the other element or intervening elements may be present. In contrast, when an element is referred to as being "directly connected" or "directly coupled" to another element, there are no intervening elements present.

In order to provide a thorough and complete description, an overview on the use of training sequences to estimate a channel response will be provided, and specific examples of DC offset, joint demodulation and transmit diversity then will be described. Embodiments of the present invention then will generally be described. A detailed description will then be provided along with a description of application of embodiments of the present invention in the above-described examples.

Overview

As was described above, time-varying multipath fading and/or other properties of the channel may make reliable receiving difficult in a wireless communication system. To reduce the effect of such time-varying channels, a sequence of training symbols, known to both the transmitter and the receiver, is often transmitted, within some or each burst of transmitted symbols, along with the data or payload symbols that are unknown to the receiver. Based on these known training symbols, the channel response corresponding to the present burst is estimated, and such a channel estimate is then used for demodulating the unknown data symbols. As an example, a conventional format of a burst of transmitted symbols in a conventional Global System for Mobile Communications (GSM)/Enhanced Data for GSM Evolution (EDGE) system is shown in Figure 1. In Figure 1, the numbers in parentheses denote the number of symbols in each segment.

A block diagram of a conventional wireless receiver is shown in Figure 2. As shown in Figure 2, the wireless receiver **200** includes a radio frequency (RF) front end **210**, an analog-to-digital (A/D) converter **220** and a sampler **230**. A demodulator **240** produces estimates of transmitted data symbols, denoted by $\hat{s}(n)$, using a channel estimate \hat{h} that is obtained by a channel estimator **250** using the known training symbols.

It is known that the performance of channel estimation techniques over the known training symbols may depend heavily on the choice of the training sequence. Even the best and the most sophisticated method of channel estimation may still result in an inaccurate channel estimate if the training sequence is not designed properly for the specific application on hand. A description of three examples where this might happen will now be provided.

First, the training-sequence selection process might not have taken into account specific implementation impairments caused by certain receiver architectures.

For example, certain receiver structures may induce a DC offset in the received signal. Channel estimation may incorporate the estimation of such a DC offset into the channel estimation. However, the selected training sequences may not work well with the estimation of a DC offset.

5 The second example may happen when receivers employ new, advanced signal processing techniques that may not have been foreseen during the development of an existing communication system. For example, a receiver may attempt to jointly demodulate a desired signal and a dominant interfering signal that may come from a base station of an adjacent cell in a cellular system. It is known that in a heavily
10 loaded cellular system, receivers are often affected by a single dominant interfering signal. Consequently, the ability of a receiver to jointly demodulate desired and interfering signals can significantly improve its performance. To perform joint demodulation, a receiver may estimate the channel for the desired signal as well as that for the interfering signal over the training period. However, the training sequence
15 used by two adjacent base stations may not necessarily have good cross correlation, which may be desirable for conventional channel estimation techniques.

 The third example is that no "good enough" training sequence may exist at all for the specific applications. This may be caused by the length of the training
20 sequence in the pre-defined burst format (26 in GSM/EDGE burst format of Figure 1) being too short and/or the size of the alphabet of the training symbols being too small (e.g. 2 in GSM). For example, the use of transmit antenna diversity (also referred to as transmit diversity), where two or more antennas are used to transmit signals simultaneously, can be an effective way of improving the multipath fading channels. In transmit diversity, if the channel experienced by one of the transmit antenna fades
25 down, the desired signal may still be recovered at the destination from the signal replica transmitted on the other antenna(s). Consequently, the probability of experiencing a complete signal outage due to multipath fading can be significantly reduced. However, to achieve this goal, the channel from each of the transmit antennas should be estimated from the received signal. To facilitate the estimation of
30 these multiple channel responses, different training sequences with good cross-correlation properties among each other may be transmitted on different antennas. However, as the number of transmit antennas increases, it may be increasingly difficult to find a proper set of training sequences.

A mathematical treatment of the three examples mentioned above, namely, DC offset, joint demodulation, and transmit diversity, within the context of an EDGE system, now will be provided. These examples will be referred to below after describing embodiments of the present invention.

- 5 As to DC offset, consider the baseband received signal with an unknown DC offset given below:

$$r(n) = \sum_{k=0}^{L-1} h_1(k) s_1(n-k) + d e^{-j3\pi/8} + w(n) \quad n = L-1, L, \dots, N-1, \quad (1)$$

- where d denotes the DC offset, $\{h_1(k)\}_{k=0}^{L-1}$ denotes the channel response to be estimated, $\{s_1(n)\}_{n=0}^{N-1}$ denotes the training symbols, $N=26$ is the length of the training symbols, and
 10 $\{w(n)\}$ denotes a white Gaussian noise process with variance σ_w^2 . Note that the term $\exp\{-j3\pi/8\}$ shown in (1) comes from the $3\pi/8$ rotation imposed on the transmitted signals of EDGE systems for the purpose of blind modulation detection. In matrix form, Equation (1) can be written as

$$\mathbf{r} = \underbrace{\begin{bmatrix} \mathbf{S}_1 & \mathbf{a} \end{bmatrix}}_{\mathbf{S}} \underbrace{\begin{bmatrix} \mathbf{h}_1 \\ d \end{bmatrix}}_{\mathbf{h}} + \mathbf{w}, \quad (2)$$

- 15 where $\mathbf{r} = (r(L-1), r(L), \dots, r(N-1))^T$, $\mathbf{h}_1 = (h_1(0), h_1(1), \dots, h_1(L))^T$, and $\mathbf{w} = (w(L), w(L+1), \dots, w(N))^T$,

$$\mathbf{S}_1 = \begin{pmatrix} s_1(L-1) & s_1(L-2) & \dots & s_1(0) \\ s_1(L) & s_1(L-1) & \dots & s_1(1) \\ \vdots & \vdots & \ddots & \vdots \\ s_1(N-1) & s_1(N-2) & \dots & s_1(N-L) \end{pmatrix} \quad \text{and} \quad \mathbf{a} = \begin{pmatrix} 1 \\ e^{-j3\pi/8} \\ e^{-j(3\pi/8)2} \\ \vdots \\ e^{-j(3\pi/8)(N-L)} \end{pmatrix}.$$

- One technique for estimating the channel \mathbf{h}_1 is to jointly estimate the vector $\mathbf{h} = (\mathbf{h}_1^T, d)^T$ given the received signal vector \mathbf{r} , and then take the first L components
 20 of the resulting estimate $\hat{\mathbf{h}}$ as an estimate of \mathbf{h}_1 . To facilitate such a joint DC and channel estimation, the training sequence $\{s_1(n)\}_{n=0}^{N-1}$ should be designed such that the columns of \mathbf{S}_1 are mutually orthogonal and are individually orthogonal to the vector \mathbf{a} , or equivalently, that the columns of the combined matrix \mathbf{S} are mutually orthogonal to each other. However, the training sequences are often designed only to
 25 make the columns of \mathbf{S}_1 to be mutually orthogonal. For instance, the columns of the

matrix S_1 that correspond to the existing training sequences used in GSM/EDGE systems are not orthogonal to the vector \mathbf{a} .

As to joint demodulation, in a receiver that jointly demodulates a desired signal and an interfering signal, the received baseband signal is modeled as

$$r(n) = \sum_{i=0}^{L-1} h_1(i) s_1(n-i) + \sum_{i=0}^{L-1} h_2(i) s_2(n-i) + w(n), \quad (3)$$

for $n = L-1, L, \dots, N-1$, where $\{h_1(i)\}_{i=0}^{L-1}$ and $\{h_2(i)\}_{i=0}^{L-1}$ denotes the channel response of the desired and interfering signal, respectively, (for simplicity, both channels are assumed to have length L), and $\{w(n)\}$ denotes a white Gaussian noise process with variance σ_w^2 . In matrix form, the received signal $r(n)$ over the training symbols can

be expressed as

$$\mathbf{r} = [\mathbf{S}_1 \ \mathbf{S}_2] \begin{bmatrix} \mathbf{h}_1 \\ \mathbf{h}_2 \end{bmatrix}, \quad (4)$$

where $\mathbf{r} = (r(L-1), r(L), \dots, r(N-1))^T$, $\mathbf{h}_1 = (h_1(0), h_1(1), \dots, h_1(L))^T$, $\mathbf{h}_2 = (h_2(0), h_2(1), \dots, h_2(L))^T$, $\mathbf{w} = (w(L-1), w(L), \dots, w(N-1))^T$,

$$\mathbf{S}_1 = \begin{pmatrix} s_1(L-1) & s_1(L-2) & \dots & s_1(0) \\ s_1(L) & s_1(L-1) & \dots & s_1(1) \\ \vdots & \vdots & \ddots & \vdots \\ s_1(N-1) & s_1(N-2) & \dots & s_1(N-L) \end{pmatrix} \quad \text{and}$$

$$\mathbf{S}_2 = \begin{pmatrix} s_2(L-1) & s_2(L-2) & \dots & s_2(0) \\ s_2(L) & s_2(L-1) & \dots & s_2(1) \\ \vdots & \vdots & \ddots & \vdots \\ s_2(N-1) & s_2(N-2) & \dots & s_2(N-L) \end{pmatrix}.$$

One technique for estimating the channels \mathbf{h}_1 and \mathbf{h}_2 is to jointly estimate both of them given the received signal vector \mathbf{r} . To facilitate such a joint channel estimation, the training sequences $\{s_1(n)\}_{n=1}^{L-1}$ and $\{s_2(n)\}_{n=1}^{L-1}$ from adjacent base stations should be designed such that the columns of $\mathbf{S} \equiv [\mathbf{S}_1, \mathbf{S}_2]$ are mutually orthogonal. However, there is no guarantee that such a property will hold in existing cellular systems, such as GSM systems.

Finally, as to transmit diversity, in a communication system employing transmit diversity with n_T transmit antenna, the received baseband signal is given by

$$r(n) = \sum_{k=1}^{n_r} \sum_{i=0}^{L-1} h_k(i) s_k(n-i) + w(n), \quad (5)$$

for $n = L-1, L, \dots, N-1$, where $\{h_k(i)\}_{i=0}^{L-1}$ denotes the channel response of length L from the transmit antenna k , and $\{w(n)\}$ denotes a white Gaussian noise process with variance σ_w^2 . In matrix form, the received signal $r(n)$ over the training symbols can be expressed as

$$\mathbf{r} = \underbrace{\begin{bmatrix} \mathbf{S}_1 & \mathbf{S}_2 & \dots & \mathbf{S}_{n_r} \end{bmatrix}}_{\mathbf{S}} \underbrace{\begin{bmatrix} \mathbf{h}_1 \\ \mathbf{h}_2 \\ \vdots \\ \mathbf{h}_{n_r} \end{bmatrix}}_{\mathbf{h}} + \mathbf{w}, \quad (6)$$

where $\mathbf{r} = (r(L-1), r(L), \dots, r(N-1))^T$, $\mathbf{h}_k = (h_k(0), h_k(1), \dots, h_k(L))^T$, $\mathbf{w} = (w(L-1), w(L), \dots, w(N-1))^T$,

$$\mathbf{S}_k = \begin{pmatrix} s_k(L-1) & s_k(L-2) & \dots & s_k(0) \\ s_k(L) & s_k(L-1) & \dots & s_k(1) \\ \vdots & \vdots & \ddots & \vdots \\ s_k(N-1) & s_k(N-2) & \dots & s_k(N-L) \end{pmatrix}.$$

- 10 One technique for estimating the channels $\{\mathbf{h}_k\}_{k=1}^{n_r}$ is to jointly estimate the vector \mathbf{h} given the received signal vector \mathbf{r} . To facilitate such a joint channel estimation, the training sequences $\{s_k\}_{k=1}^{n_r}$ should be designed such that the columns of $\mathbf{S} = [\mathbf{S}_1, \mathbf{S}_2, \dots, \mathbf{S}_{n_r}]$ are mutually orthogonal. However, if the length of the training sequences and/or the size of their alphabet are not large enough, there may not exist
- 15 any set of training sequences with, or close to having, such a property.

Note that the received signals in matrix form (Equations (2) and (6)) for both the example of DC offset and the examples of transmit diversity (or the combination of both) can be formulated in the following general form:

$$\mathbf{r} = \mathbf{S}\mathbf{h} + \mathbf{w}, \quad (7)$$

- 20 where \mathbf{w} is a Gaussian-distributed noise vector. In channel estimation, it is desired to estimate the vector \mathbf{h} given the received signal \mathbf{r} . Embodiments of the invention can improve the estimation of \mathbf{h} over conventional estimation methods when the columns of the matrix \mathbf{S} are not mutually orthogonal to each other. For simplicity, the noise vector \mathbf{w} is assumed to be white in the following descriptions. However, one or

ordinary skill in the art can generalize to the case when the noise vector \mathbf{w} has non-white statistics.

General Description

5 Embodiments of the present invention can provide systems, methods and receivers that systematically add bias to the channel estimate that will result in a smaller overall estimation error than, say, the conventional channel estimation methods, which may compute a non-biased channel estimate. It can be shown that by trading off bias with the statistical error caused by the additive noise process,
10 according to the characteristics of the matrix \mathbf{S} in some embodiments of the invention, overall estimation error can be significantly reduced. More precisely, consider the overall mean squared estimation error, which can be expressed as sum of two terms:

$$\text{tr}[\mathbf{E}(\mathbf{h} - \hat{\mathbf{h}})(\mathbf{h} - \hat{\mathbf{h}})^H] = \underbrace{\text{tr}[(\mathbf{h} - \mathbf{E}\hat{\mathbf{h}})(\mathbf{h} - \mathbf{E}\hat{\mathbf{h}})^H]}_{\text{Bias}} + \underbrace{\text{tr}[\mathbf{E}(\hat{\mathbf{h}} - \mathbf{E}\hat{\mathbf{h}})(\hat{\mathbf{h}} - \mathbf{E}\hat{\mathbf{h}})^H]}_{\text{Statistical Error}} \quad (8)$$

15 Some embodiments of the invention can compute an estimate $\hat{\mathbf{h}}$ which can provide a good tradeoff between the two terms in (8).

Figure 3 is a block diagram of systems, methods and receivers for estimating a channel response from training symbols that are received over a channel according to some embodiments of the present invention. These embodiments may be used, for example, in the channel estimator **250** of Figure 2. As shown in Figure 3 at Block **310**, an initial channel estimating circuit is configured to determine an initial channel estimate $\hat{\mathbf{h}}_0$ from the training symbols \mathbf{r} that are received. Conventional channel estimation techniques may be used. At Block **320**, a bias applying circuit is configured to apply a bias to the initial channel estimate $\hat{\mathbf{h}}_0$, to obtain a biased channel estimate $\hat{\mathbf{h}}$.
25

Figure 4 is a flowchart of operations that may be performed to estimate a channel response according to other embodiments of the present invention. These operations may be used, for example, by the channel estimator **250** of Figure 2.

As shown in Figure 4 at Block **410**, an initial channel estimate is determined from the training symbols that are received, for example using conventional techniques. At Block **420**, bias is added to obtain a biased channel estimate. At Block **430**, a test is made as to whether additional improvement is desired and, if so,
30

operations at Block **420** may be repeatedly performed until a desired improvement is obtained. Once obtained, at Block **440**, the biased channel estimate is used to demodulate a signal that is received over the channel.

A common technique for estimating \mathbf{h} in (7) is the least squares (LS) channel estimation. The LS estimate of \mathbf{h} given \mathbf{r} is :

$$\hat{\mathbf{h}}_{LS} = (\mathbf{S}^H \mathbf{S})^{-1} \mathbf{S}^H \mathbf{r}, \quad (9)$$

which is known to be an unbiased estimate of the true channel \mathbf{h} , i.e. $E \hat{\mathbf{h}}_{LS} = \mathbf{h}$. In effect, such an estimate constrains the bias, as shown in the first term in (8), to be zero and then minimizes the statistical error, as shown in the second term in (8), for the given matrix \mathbf{S} . However, when the matrix \mathbf{S} does not have mutually orthogonal columns, by allowing a non-zero bias, the statistical error can be much further reduced.

Figure 5 is a block diagram of other embodiments of the present invention. In these embodiments, the receiver first performs an initial unbiased channel estimation, for example the LS channel estimation described in (9), at Block **510**. Then, at Block **520**, a circuit is configured to transform the initial channel estimate. For example, the resulting channel estimate $\hat{\mathbf{h}}_o$ is transformed according to a pre-computed orthogonal matrix Ψ at Block **520**. In some embodiments, the matrix Ψ has the eigenvectors of (a) the matrix $\mathbf{S}^H \mathbf{S}$ or (b) the matrix $(\hat{\mathbf{h}}_o \hat{\mathbf{h}}_o^H - \sigma_w^2 (\mathbf{S}^H \mathbf{S})^{-1})$ as columns, where the noise power σ_w^2 may be obtained from the initial channel estimation. At Block **530**, a circuit is configured to perform an operation on the transformed channel estimate to change at least one of its values. In some embodiments, at Block **530**, all but K taps (where $K < L$) of the transformed channel estimate $\hat{\mathbf{c}}_o$ are set to zero according to certain selection criteria. At Block **540**, a circuit is configured to inverse transform the transformed initial channel estimate on which the operation was performed. For example, the channel estimate $\hat{\mathbf{h}}$ (after the first iteration) is obtained by inverse transforming the resulting estimate $\hat{\mathbf{c}}$ at Block **540**.

In some embodiments, the number K may be predetermined according to the distribution of eigenvalues of the matrix $\mathbf{S}^H \mathbf{S}$ and stored in the receivers. In these embodiments, all taps of $\hat{\mathbf{c}}_o$ but those corresponding to the K largest eigenvalues of $\mathbf{S}^H \mathbf{S}$ are set to zero. In other embodiments, K may be selected to be the number of positive eigenvalues of the matrix $(\hat{\mathbf{h}}_o \hat{\mathbf{h}}_o^H - \sigma_w^2 (\mathbf{S}^H \mathbf{S})^{-1})$. These embodiments can be

equivalent to selecting K by comparing each individual element of the transformed channel estimate $\hat{\mathbf{c}}_o$ to at least one threshold and then setting K to be the number of elements that are greater than the corresponding thresholds. More precisely, the i -th threshold for the i -th tap of $\hat{\mathbf{c}}_o$ to compare may be given by

$$5 \quad \gamma_i = \frac{\sigma_w^2}{[\Psi^H (S^H S) \Psi]_{ii}}$$

where $[A]_{ii}$ denotes the i -th diagonal element of the matrix A , and the columns of Ψ here are the eigenvectors of the matrix $(\hat{\mathbf{h}}_o \hat{\mathbf{h}}_o^H - \sigma_w^2 (S^H S)^{-1})$.

For example, when the LS channel estimation in (9) is employed as the initial channel estimation $\hat{\mathbf{h}}_o$, the new channel estimate $\hat{\mathbf{h}}$ is given by:

$$10 \quad \hat{\mathbf{h}} = \Psi_K \Psi_K^H \hat{\mathbf{h}}_o = \Psi_K \Psi_K^H (S^H S)^{-1} S^H \mathbf{r}. \quad (10)$$

where Ψ_K denotes a L by K matrix, whose columns are mutually orthogonal to each other and may be chosen as (a) the K eigenvectors of the matrix $S^H S$ with the largest eigenvalues, or as (b) those eigenvectors of the matrix $(\hat{\mathbf{h}}_o \hat{\mathbf{h}}_o^H - \sigma_w^2 (S^H S)^{-1})$ that have positive eigenvalues. If $K = L$, i.e. all taps of the transformed channel estimate $\hat{\mathbf{c}}_o$ are preserved, the estimate in (10) reduces to the ordinary LS channel estimate described in (9).

Figure 5 also includes an optional feedback loop 550 according to other embodiments of the invention, where the channel estimate $\hat{\mathbf{h}}$ from the first iteration is used to compute a new transform Ψ , which is then applied to the initial channel estimate $\hat{\mathbf{h}}_o$ again. A new value of K or a set of new thresholds to which the elements of $\hat{\mathbf{c}}_o$ are compared may also be updated. In some embodiments, the new transform is computed using the eigenvectors of the matrix $(\hat{\mathbf{h}} \hat{\mathbf{h}}^H - \sigma_w^2 (S^H S)^{-1})$ as the columns. More than two iterations may be executed through the feedback loop to further improve the final channel estimate $\hat{\mathbf{h}}$.

Figure 6 is a flowchart of operations that may be performed according to other embodiments of the present invention. As shown in Figure 6, an initial unbiased channel estimate is computed at Block 610. At Block 620, the initial channel estimate is converted into a transformed channel estimate. At Block 630, a number of taps K to keep in the transformed channel estimate is selected if this number is not predetermined. At Block 640, all but K -taps of the transformed channel estimate are

set to 0. At Block 650, the inverse transform is applied to the resulting bias transformed channel estimate. At Block 660, if additional improvement is desired, then at Block 670 the new transform and new thresholds are computed if desired.

In still other embodiments of the invention, at Blocks 530 or 640, K -taps are set to 0 and the remaining taps are kept. In these embodiments, the new channel estimate may be subtracted from the initial channel estimate rather than being added thereto at Block 330 of Figure 3.

Detailed Explanation

10 A detailed explanation of introducing intentional bias into an initial channel estimate in order to reduce the overall estimation error, according to some embodiments of the present invention, now will be provided. It will be shown that when no *a priori* information about the true channel \mathbf{h} is known, the channel estimate in (10) that resulted from the choice of the columns of Ψ_K being

- 15 (a) the K eigenvectors of the matrix $\mathbf{S}^H \mathbf{S}$ with the largest eigenvalues or
- (b) the eigenvectors of the matrix $(\hat{\mathbf{h}}_0 \hat{\mathbf{h}}_0^H - \sigma_w^2 (\mathbf{S}^H \mathbf{S})^{-1})$ with positive eigenvalues,

gives a good tradeoff between the bias and the statistical error as shown in (8) for any $K < L$.

20 Let Ψ be the unitary matrix in Figure 5, let $\mathbf{c} = (c_1, c_2, \dots, c_L)^T = \Psi^H \mathbf{h}$ (or equivalently $\mathbf{h} = \Psi \mathbf{c}$) and let $\hat{\mathbf{c}} = (\hat{c}_1, \hat{c}_2, \dots, \hat{c}_L)^T = \Psi^H \hat{\mathbf{h}}$. Also let $I = \{i_1, i_2, \dots, i_K\}$ be a set of integers such that $1 \leq i_1 < i_2 < \dots < i_K \leq N$, which will be determined later.

Corresponding to any given I , let $\mathbf{J}(I)$ (or just \mathbf{J}) denote an L by K matrix with the (i_k, k) element being one, where $k = 1, 2, \dots, K$, and with all other elements being zero.

25 Note that multiplying a channel estimate $\hat{\mathbf{c}}_0$ by $\mathbf{J}\mathbf{J}^H$ is the same as setting all taps whose indices are not in I to zero.

A formula of the overall estimation error defined in (8) for embodiments of the invention depicted in Figure 5 will be described. Then, the estimation error can be minimized over integer sets I (or equivalently, \mathbf{J} in the format described above) and unitary transforms Ψ , under the following constraints on the estimate $\hat{\mathbf{h}}$: (1) $\hat{c}_j = 0$ for all $j \notin I$, and (2) $\mathbf{E}\hat{\mathbf{h}}_0 = \mathbf{h}$ for all $j \in I$.

From (8), the bias of $\hat{\mathbf{h}}$ (i.e. the first term in (8)) is given by

$$\begin{aligned}
 \Delta_{\epsilon}^2 &= \text{tr}[(\mathbf{h} - E\hat{\mathbf{h}})(\mathbf{h} - E\hat{\mathbf{h}})^H] \\
 &= \text{tr}[\Psi(\mathbf{c} - E\hat{\mathbf{c}})(\mathbf{c} - E\hat{\mathbf{c}})^H \Psi^H] \\
 &= \text{tr}[(\mathbf{c} - E\hat{\mathbf{c}})(\mathbf{c} - E\hat{\mathbf{c}})^H] \\
 &= \|\mathbf{c}\|^2 - \mathbf{c}^H(E\hat{\mathbf{c}}) - (E\hat{\mathbf{c}})^H\mathbf{c} + \|E\hat{\mathbf{c}}\|^2 \\
 &= \|\mathbf{h}\|^2 - \|\mathbf{J}^H \mathbf{c}\|^2 \\
 &= \|\mathbf{h}\|^2 - \text{tr}[\mathbf{J}^H \Psi^H \mathbf{h} \mathbf{h}^H \Psi \mathbf{J}]
 \end{aligned}$$

where the fifth equality follows from the fact that $E\hat{\mathbf{c}} = E\mathbf{J}\mathbf{J}^H\hat{\mathbf{c}}_o = \mathbf{J}\mathbf{J}^H(E\hat{\mathbf{c}}_o) = \mathbf{J}\mathbf{J}^H\mathbf{c}$ since $\hat{\mathbf{c}}_o$ is an unbiased estimate of \mathbf{c} . By substituting $\hat{\mathbf{h}} = \Psi\mathbf{J}\mathbf{J}^H\Psi^H\hat{\mathbf{h}}_o$ in the second term of (8), the statistical error is obtained as

$$\begin{aligned}
 \sigma_{\epsilon}^2 &= \text{tr}[E(\hat{\mathbf{h}} - E\hat{\mathbf{h}})(\hat{\mathbf{h}} - E\hat{\mathbf{h}})^H] \\
 &= \text{tr}[E\Psi\mathbf{J}\mathbf{J}^H\Psi^H(\hat{\mathbf{h}}_o - E\hat{\mathbf{h}}_o)(\hat{\mathbf{h}}_o - E\hat{\mathbf{h}}_o)^H\Psi\mathbf{J}\mathbf{J}^H\Psi^H] \\
 &= \text{tr}[E\mathbf{J}^H\Psi^H(\hat{\mathbf{h}}_o - E\hat{\mathbf{h}}_o)(\hat{\mathbf{h}}_o - E\hat{\mathbf{h}}_o)^H\Psi\mathbf{J}] \\
 &= \text{tr}[E\mathbf{J}^H\Psi^H(\hat{\mathbf{h}}_o - \mathbf{h})(\hat{\mathbf{h}}_o - \mathbf{h})^H\Psi\mathbf{J}] \\
 &\geq \text{tr}[E\mathbf{J}^H\Psi^H\left(-\frac{\partial^2 \ln p(\mathbf{r}|\mathbf{h})}{\partial \mathbf{h}^2}\right)^{-1}\Psi\mathbf{J}] \\
 &= \sigma_w^2 \text{tr}[\mathbf{J}^H\Psi^H(S^H S)^{-1}\Psi\mathbf{J}]
 \end{aligned}$$

where the fourth equality above follows from the fact that $\hat{\mathbf{h}}_o$ is an unbiased estimate of \mathbf{h} , and where the inequality follows from the Cramer-Rao bound for general unbiased estimates. This inequality becomes equality when the ordinary LS estimation is used for computing the initial channel estimate. In other words, to minimize the statistical error, one should choose

$$\hat{\mathbf{h}}_o = (S^H S)^{-1} S^H \mathbf{r}, \quad (11)$$

or in other words,

$$\hat{\mathbf{h}} = \Psi\mathbf{J}\mathbf{J}^H\Psi^H\hat{\mathbf{h}}_o = \Psi\mathbf{J}\mathbf{J}^H\Psi^H(S^H S)^{-1} S^H \mathbf{r}. \quad (12)$$

Noting that $\Psi_K = \Psi\mathbf{J}$, the proposed estimate shown in (10) is obtained.

Next, a choice for Ψ and \mathbf{J} according to some embodiments of the present invention will be described. By summing the contribution from the bias and the statistical error, the overall estimation error becomes

$$\begin{aligned}
 \text{tr}[E(\mathbf{h} - \hat{\mathbf{h}})(\mathbf{h} - \hat{\mathbf{h}})^H] &= \Delta_{\epsilon}^2 + \sigma_{\epsilon}^2 \\
 &= \|\mathbf{h}\|^2 - \text{tr}[\mathbf{J}^H\Psi^H(\mathbf{h}\mathbf{h}^H - \sigma_w^2(S^H S)^{-1})\Psi\mathbf{J}] \\
 &= \|\mathbf{h}\|^2 - \sum_{k=1}^K \phi_k^H(\mathbf{h}\mathbf{h}^H - \sigma_w^2(S^H S)^{-1})\phi_k
 \end{aligned} \quad (13)$$

where ϕ_i denotes the i th column of Ψ . From (13), it appears that the columns of the best matrix Ψ should be the one with the eigenvectors of the matrix $(\mathbf{h}\mathbf{h}^H - \sigma_w^2(\mathbf{S}^H\mathbf{S})^{-1})$, and a set of integers $\{i_k\}_{k=1}^K$ should be those indices that corresponds to the positive eigenvalues of $(\mathbf{h}\mathbf{h}^H - \sigma_w^2(\mathbf{S}^H\mathbf{S})^{-1})$, where K here represents the number of eigenvalues of $(\mathbf{h}\mathbf{h}^H - \sigma_w^2(\mathbf{S}^H\mathbf{S})^{-1})$ that are positive.

However, since $\mathbf{h}\mathbf{h}^H$ is not known a priori, one may use $\hat{\mathbf{h}}_o\hat{\mathbf{h}}_o^H$ as an estimate of $\mathbf{h}\mathbf{h}^H$, which leads to (b) the transform $(\hat{\mathbf{h}}_o\hat{\mathbf{h}}_o^H - \sigma_w^2(\mathbf{S}^H\mathbf{S})^{-1})$. Alternatively, one may optimize Ψ in a minimax sense (i.e. for the worse channel realization) in the first iteration:

$$\begin{aligned} \Psi &= \arg \min_{\Psi: \Psi^H \Psi = I} \max_{\mathbf{h}: \|\mathbf{h}\|^2 \leq P} \left[\|\mathbf{h}\|^2 - \sum_{k=1}^K \phi_{i_k}^H (\mathbf{h}\mathbf{h}^H - \sigma_w^2(\mathbf{S}^H\mathbf{S})^{-1}) \phi_{i_k} \right] \\ &= \arg \min_{\Psi: \Psi^H \Psi = I} \left[\sigma_w^2 \sum_{k=1}^K \phi_{i_k}^H (\mathbf{S}^H\mathbf{S})^{-1} \phi_{i_k} \right] \end{aligned} \quad (14)$$

where P is any positive constant indicating the maximum strength of the channel response \mathbf{h} under consideration. It appears from (14) that the columns of the transform Ψ should be chosen as the eigenvectors of (b) the matrix $\mathbf{S}^H\mathbf{S}$, and the set of integers $\{i_k\}_{k=1}^K$ should be those indices corresponding to the largest eigenvalues of $\mathbf{S}^H\mathbf{S}$. One may also think of this choice of Ψ as designing for the low signal-to-noise situation where σ_w^2 is large.

Embodiments of the invention may significantly improve performance of receivers that have a DC offset problem, such as the homodyne receivers typically used for EDGE. Figure 7 shows the gains that may be achieved by an embodiment of the invention with one iteration, as K varies, in mean-squared estimation error (in dB) over the LS channel estimation, which is also equal to the Cramer-Rao lower bound on any unbiased channel estimation techniques. The eigenvectors of the matrix $\mathbf{S}^H\mathbf{S}$ are used as the columns of the transform Ψ . The GSM training sequence TS5 and a joint estimation of a 7-tap baseband channel and 1-tap DC at an C/N ratio of 10dB are assumed in this figure, which shows that up to 2dB gain in mean squared estimation error can be achieved by reducing K from $L = 8$ to 6.

Figure 8 shows the corresponding gain in mean-squared estimation error in the case of joint channel estimation of two 7-tap channels for transmit diversity. A C/N ratio of 18dB and the GSM training sequences TS0 and TS5 and are assumed in the

figure. It is shown that up to 11dB in mean-squared channel estimation can be achieved by reducing K from $L = 14$ to 11.

In the drawings and specification, there have been disclosed embodiments of the invention and, although specific terms are employed, they are used in a generic and descriptive sense only and not for purposes of limitation, the scope of the invention being set forth in the following claims.